MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RE-45

NONLINEAR ESTIMATION WITH QUANTIZED MEASUREMENTS: PCM PREDICTIVE QUANTIZATION, AND DATA COMPRESSION

by

Renwick E. Curry

August 1968

EXPERIMENTAL ASTRONOMY LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE 39, MASSACHUSETTS

RE-45

NONLINEAR ESTIMATION WITH QUANTIZED MEASUREMENTS: PCM, PREDICTIVE QUANTIZATION, AND DATA COMPRESSION

by

Renwick E. Curry

August, 1968

EXPERIMENTAL ASTRONOMY LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS, 02139

Approved: W. Markey

Director
Experimental Astronomy Laboratory

ACKNOWLEDGEMENT

The author wishes to thank T. Fine and H. Gish for their helpful discussions on the predictive quantization problem.

This report was sponsored by the National Aeronautics and Space Administration through research grant NsG 254-62.

The publication of this report does not constitute approval by the National Aeronautics and Space Administration of the findings or the conclusions contained therein. It is published only for the exchange and stimulation of ideas.

This work is based on a portion of the author's Ph.D. dissertation submitted to the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, Mass.

NONLINEAR ESTIMATION WITH QUANTIZED MEASUREMENTS:
PCM, PREDICTIVE QUANTIZATION, AND DATA COMPRESSION

Renwick E. Curry

ABSTRACT

Statistics conditioned on quantized measurements are considered in the general case. These results are specialized to Gaussian parameters and then extended to discrete time linear systems. The conditional mean of the system's state vector may be found by passing the conditional mean of the measurement history through the Kalman Filter that would be used had the measurements been linear. Repetitive use of Bayes' Rule is not required. Because the implementation of this result requires lengthy numerical quadrature, two approximations are considered: the first is a power series expansion of the probability density function; the second is a discrete time version of a previously proposed algorithm which assumes that the conditional distribution is normal. Both algorithms may be used with any memory length on stationary or nonstationary data. The two algorithms are applied to the noiseless channel versions of the PCM, predictive quantization, and predictive-comparison data compression systems; ensemble average performance estimates of the nonlinear filters are derived. Simulation results show that the performance estimates are quite accurate for most of the cases tested.

PRECEDING PAGE BLANK NOT FILMED.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

I. INTRODUCTION AND SUMMARY

The increasing demand on existing digital facilities (e.g., communication channels, data storage) can be alleviated by representing the same amount of information with fewer bits at the expense of more sophisticated data processing. The current interest in this area [see 1] goes by the various names of "redundancy reduction", "data compression", and others. Common to the great majority of these approaches is the problem of computing estimates from quantized data. This paper considers the analysis and implementation of nonlinear estimation with quantized measurements, and these techniques are applied to three types of digital systems: PCM, predictive quantization, and predictive-comparison data compression.

Relevant Work

Much work has been done in the area of linear filtering of quantized measurements [2,3,4], and some has been done on the general nonlinear filtering problem [e.g., 5,6,7]. Unfortunately, these latter treatments require either the repetitive use of Bayes' Rule, or the power series expansion of the nonlinear measurement function, a technique which is not applicable to the quantizer's staircase input-output graph. To the best of the author's knowledge, only Meier, Korsak, and Larson [8] have specifically considered nonlinear estimation with quantized measurements. They derive some nonlinear estimates including the conditional mean and covariance of a scalar state based on one quantized measurement. The a priori distribution for both scalars is assumed Gaussian, and they indicate that subsequent estimates must be found by repetitive use of Bayes' Rule since the posterior distribution is no longer normal.

The PCM problem is considered by Ruchkin [2], Steiglitz [3], and Kellog [4] who use linear filtering to reconstruct the system's input. Fine [9] gives a theoretical treatment of optimum digital systems with an example of predictive quantization (feedback around a binary quantizer). Bello, Lincoln, and Gish [10] have computed Fine's nonlinear feedback function by Monte Carlo techniques and give some simulation results. Gish [11], O'Neal [12], and Irwin and O'Neal [13] consider the problem of designing linear feedback functions. Predictive-comparison data compression systems are in the class of predictive quantization systems. In [14] Davisson treats the optimum linear feedback operation; in [15] he considers an adaptive system of the same type.

Summary

Linear estimation uses the quantizer's staircase input-output graph. Nonlinear estimation uses the information that the quantized measurements z lie in an hypercube (say) A. In Section II it is shown that moments conditioned on quantized measurements can be computed in two steps: 1) Find the expectation conditioned on a measurement z (the usual estimation problem); 2) Average this function of z conditioned on $z \in A$. These results are specialized to Gaussian distributions and then extended to linear dynamical systems: the conditional mean of the state vector is found by passing the conditional mean of the measurement history through the Kalman Filter that would be used had the measurements been linear. Thus the conditional mean can be computed without using Bayes' Rule.

Section III considers two approximate nonlinear filters to compute the conditional mean of stationary and nonstationary data. The first involves a power series expansion to find the conditional probability density function. The second is a recursive computation which approximates the conditional distribution just prior to a quantized measurement by a normal distribution. Section IV applies these approximations to the noiseless channel versions of the PCM, predictive quantization, and predictive-comparison data compression systems. The first two systems are designed to minimize the mean square reconstruction error, whereas the data compression system is designed to

minimize (approximately) the average number of samples sent to the receiver. Estimates of the ensemble performance are derived for each system. This is extremely important in design work since it allows evaluation of system performance without Monte Carlo simulation.

The three systems have been simulated on a digital computer, and the actual performance is compared to the ensemble estimates in Section V; agreement is quite good for most of the cases tested. Section VI closes with a summary of the major conclusions and limitations of the results.

PRECEDING PAGE BLANK NOT FILMED.

II. CONDITIONAL STATISTICS

Preliminaries

Let x be an n-component parameter vector and z an m-component measurement vector. They are related through a (perhaps nonlinear) measurement equation that may or may not include noisy observations. It is assumed that the joint probability density function of x and z, $p_{x,z}(\xi,\zeta)$, exists and is known. Let the individual components of z be quantized and thus $z\epsilon A$ implies $\{a^i \leq z^i < b^i,\ i=1,..m\}$ (The majority of the following results do not depend on the fact that A is an hypercube.)

Conditional Expectations

It is well known [e.g., 16] that if $\mathcal{F}_1$ and $\mathcal{F}_2$ are Borel fields and $\mathcal{F}_1 \subset \mathcal{F}_2$, then for the random variable y

$$\mathcal{E}\left[\mathcal{E}(y|\mathcal{F}_2)|\mathcal{F}_1\right] = \mathcal{E}(y|\mathcal{F}_1) \tag{1}$$

For the purposes of estimation with quantized measurements let y be f(x), $\mathcal{F}_1$ be $z\epsilon A$, $\mathcal{F}_2$ be z. The equality (1) is reversed, i.e., the unknown quantity becomes the known quantity, with the result that

$$\mathcal{E}[f(x)|z\epsilon A] = \mathcal{E}[\mathcal{E}(f(x)|z)|z\epsilon A] \tag{2}$$

Observe that the expectation of f(x) conditioned on quantized measurements can be performed in two steps: 1) Find $\mathcal{E}(f(x)|z)$, this is the usual goal of estimation with unquantized measurements; 2) Find the expectation of $\mathcal{E}(f(x)|z)$ conditioned on $z\epsilon A$. For step 2 one must use the probability density function for z conditioned on $z\epsilon A$.

$$p_{z|z\epsilon A}(\zeta) = \begin{cases} \dfrac{p_z(\zeta)}{P(z\epsilon A)} = \dfrac{p_z(\zeta)}{\int_A p_z(\zeta)d\zeta}, & \zeta\epsilon A \\ 0, & \zeta \not\epsilon A \end{cases} \tag{3}$$

where $p_z(\zeta)$ is the a priori probability density of the measurement vector.

Gaussian Parameters

Let the distribution of the parameter vector x be normal with mean $\bar{x}$ and covariance M: $N[\bar{x},M]$. Let the measurements be linearly related to x with additive independent observation noise.

$$z = Hx+v \tag{4}$$

where H is an mxn matrix and v is the noise vector $N[0,R]$. To find the mean of x conditioned on quantized measurements let $f(x) = x$. The results of executing the first step in the procedure are well known.

$$\mathcal{E}(x|z) = \bar{x}+K[z-H\bar{x}] \tag{5}$$

where

$$K = MH^T(HMH^T+R)^{-1} \tag{6}$$

The second step in the procedure is to take the expectation of (5) conditioned on $z \epsilon A$ to find the mean of x conditioned on $z \epsilon A$.

$$\mathcal{E}(x|z\epsilon A) = \bar{x} + K[\mathcal{E}(z|z\epsilon A)-H\bar{x}] \tag{7}$$

It is easily shown [17] that the covariance in the estimate conditioned on $z \epsilon A$ is given by

$$\mathrm{cov}(x|z\epsilon A) = P + K\ \mathrm{cov}(z|z\epsilon A)K^T \tag{8}$$

where P is the covariance of x that would be obtained had the measurements been linear:

$$P = M-MH^T(HMH^T+R)^{-1}HM \tag{9}$$

Both the conditional mean and covariance reduce to the proper form when the measurements are unquantized since the probability density function of z conditioned on $z \epsilon A$ approaches an impulse. Note the importance in the nonlinear estimate (7) of the gain matrix K which is used to weight linear measurements. Interestingly enough, (8) shows that quantization increases the (minimum) variance as though it were uncertainty added <u>after</u> a linear measurement had been processed. This is in sharp distinction to the point of view that treats quantization as observation noise added <u>before</u> the linear measurement.

<u>Extension to Linear Dynamic Systems</u>

Assume that a Gauss-Markov process is described by the following equations:

$$x_{i+1} = \Phi_i x_i + w_i \qquad i=0,\ldots,K \tag{10}$$

$$z_i = H_i x_i + v_i \qquad i=1,\ldots,K \tag{11}$$

$$x_0 \text{ is } N[\bar{x}_0, P_0], \tag{12}$$

$$\mathcal{E}(w_i)=0,\ \mathcal{E}(w_i w_j^T) = Q_i \delta_{ij} \tag{13}$$

$$\mathcal{E}(v_i)=0,\ \mathcal{E}(v_i v_j^T) = R_i \delta_{ij} \tag{14}$$

$$\mathcal{E}(w_i v_j^T) = \mathcal{E}(w_i x_0^T) = \mathcal{E}(v_i x_0^T) = 0 \tag{15}$$

where

x_i = system state vector at time t_i

Φ_i = system transition matrix from time t_i to t_{i+1}

w_i = Gaussian process noise at time t_i

z_i = measurement vector at time t_i

H_i = measurement matrix at time t_i

v_i = Gaussian observation noise at time t_i

Filtering, as used here, means the determination of the conditional mean of a state vector using present and past quantized measurements; prediction is the determination of the conditional mean of a state vector using only past quantized measurements; smoothing is the determination of the conditional mean of a state vector using past, present, and future quantized measurements. (A "future" measurement occurs at a time later than the state vector in question.)

The solution to these three problems can be found by using the solution to the Gaussian parameter estimation problem considered earlier since the a priori distribution of all random variables is normal.

1) Let x be the state vector(s) under consideration. Let z be the collection of all measurement vectors, and let A be the region in which they fall.

2) Compute the mean of z conditioned on $z \epsilon A$.

3) Solve (7) for the mean of the state vector(s) conditioned on $z \epsilon A$.

Remarks

a) The prediction solution is directly obtained from the filtering solution because of the independence of the process noise vectors $\{w_i\}$, i.e. from (10)

$$\mathcal{E}(x_{i+n} | z_i \epsilon A_i, z_{i-1} \epsilon A_{i-1} \ldots) =$$

$$\Phi_{i+n-1} \Phi_{i+n-2} \ldots \Phi_i \mathcal{E}(x_i | z_i \epsilon A_i, z_{i-1} \epsilon A_{i-1}, \ldots)$$

$$n > 0$$

Only the conditional mean of the state vector is considered in the sequel since the mean of a linear function of the state is the same linear function of the mean.

b) Eq. (7) now represents the "batch processing" solution to the estimation problem. These equations may be solved recursively via the Kalman Filter using $\mathcal{E}(z_i|z\epsilon A)$ as the filter input at time t_i. (Additional computations on these filter inputs are required for the smoothing problem [18].)

c) This formulation yields the conditional mean of the state vector for nonstationary data and arbitrary quantization schemes. Note that repeated use of Bayes' Rule is not required.

d) If z is a scalar $N[\bar{z},\sigma_z^2]$, and A is defined by $\{z|a\leq z<b\}$, then $\mathcal{E}(z|z\epsilon A)$ and $\mathrm{cov}(z|z\epsilon A)$ are given by

$$\mathcal{E}(z|\,z\epsilon A) = \bar{z} + \frac{\sigma_z}{P(z\epsilon A)}\left[\frac{e^{-\frac{1}{2}\left(\frac{a-\bar{z}}{\sigma_z}\right)^2}}{\sqrt{2\pi}} - \frac{e^{-\frac{1}{2}\left(\frac{b-\bar{z}}{\sigma_z}\right)^2}}{\sqrt{2\pi}}\right] \tag{16}$$

$$\mathrm{cov}(z|z\epsilon A) = \sigma_z^2\left\{1 + \left(\frac{a-\bar{z}}{\sigma_z}\right)\frac{e^{-\frac{1}{2}\left(\frac{a-\bar{z}}{\sigma_z}\right)^2}}{\sqrt{2\pi}\,P(z\epsilon A)} - \left(\frac{b-\bar{z}}{\sigma_z}\right)\frac{e^{-\frac{1}{2}\left(\frac{b-\bar{z}}{\sigma_z}\right)^2}}{\sqrt{2\pi}\,P(z\epsilon A)}\right. \tag{17}$$

$$\left. - \frac{1}{P(z\epsilon A)^2}\left[\frac{e^{-\frac{1}{2}\left(\frac{a-b}{\sigma_z}\right)^2}}{\sqrt{2\pi}} - \frac{e^{-\frac{1}{2}\left(\frac{b-\bar{z}}{\sigma_z}\right)^2}}{\sqrt{2\pi}}\right]^2\right\}$$

where

$$P(z\epsilon A) = \int_{\frac{a-\bar{z}}{\sigma_z}}^{\frac{b-\bar{z}}{\sigma_z}} \frac{e^{-\frac{1}{2}v^2}}{\sqrt{2\pi}}\,dv \tag{18}$$

For two or more components numerical quadrature is required to obtain a precise computation of $\mathcal{E}(z|z\epsilon A)$. Despite these difficulties, this approach does provide one common point of departure for design purposes, and approximations can be made which depend on the specific method of quantization.

PRECEDING PAGE BLANK NOT FILMED.

III. APPROXIMATE NONLINEAR ESTIMATION

Power Series Expansion

Here we describe a power series method to approximate the mean and covariance of an m-component zero mean Gaussian vector conditioned on quantized measurements. Its use is restricted to cases where the quantum-interval-to-standard-deviation ratio is small since fourth and higher order terms are neglected.

Let the vector inequality define the quantum region A

$$a \leq z < b \tag{19}$$

i.e. $\{a^i \leq z^i < b^i,\ i=1,\ldots,m\}$. The geometric center of A is the vector γ

$$\gamma = \frac{1}{2}(b+a) \tag{20}$$

and the vector of quantum interval halfwidths is $\{\alpha^i\}$.

$$\alpha = \frac{1}{2}(b-a) \tag{21}$$

The Gaussian probability density function is expanded in a power series about γ and fourth and higher order terms are neglected. The details of this straightforward but lengthy procedure have been carried out in [17] with the result that

$$\mathcal{E}(z|z\epsilon A) \approx (I - C\Gamma^{-1})\gamma \tag{22}$$

$$\mathrm{cov}(z|z\epsilon A) \approx C = \left\{ \frac{(\alpha^i)^2}{3} \delta_{ij} \right\} \tag{23}$$

where

$$\Gamma = \mathcal{E}(zz^T)$$

The conditional mean is given by the geometric center of A plus a second order correction term. This approximation, given by (22), looks deceptively linear, but it is nonlinear because both γ and α are part of the measurement. Both (22) and (23) may be used in (7) and (8) to give estimates of the mean and covariance of x conditioned on quantized measurements.

The Gaussian Fit Algorithm

The Gaussian Fit Algorithm is the present author's name for the discrete time version of a suboptimal filtering algorithm suggested independently by Jazwinski [19], Bass and Schwartz [20], and, apparently, by Fisher [21]. Here we present some heuristic justification for the technique and derive (in the Appendix) an ensemble average performance estimate even for nonstationary data.

Consider the system described by (10-15) and assume that

A1: The conditional distribution of the state just prior to the ith measurement is $N[\hat{x}_{i|i-1}, M_i]$

Then we know from Section II and (10-15) that

$$\hat{x}_{i|i}=\hat{x}_{i|i-1}+K_i[\mathcal{E}(z_i|z_i \epsilon A_i)-H_i\hat{x}_{i|i-1}] \tag{24}$$

$$K_i=M_iH_i(H_iM_iH_i^T+R_i)^{-1} \tag{25}$$

$$P_i=M_i-M_iH_i^T(H_iM_iH_i^T+R_i)^{-1}H_iM_i \tag{26}$$

$$E_i=P_i+K_i\text{cov}(z_i|z_i \epsilon A_i)K_i^T \tag{27}$$

$$\hat{x}_{i+1|i}=\Phi_i\hat{x}_{i|i} \tag{28}$$

$$M_{i+1}=\Phi_iE_i\Phi_i^T+Q_i \tag{29}$$

where the newly introduced symbols are

$\hat{x}_{i|i}$ = conditional mean (under A1) of x_i given quantized measurements up to and including t_i.

$\hat{x}_{i|i-1}$ = conditional mean (under A1) of x_i given quantized measurements up to and including t_{i-1}

A_i = quantum region for z_i

M_i = conditional covariance (under A1) of x_i given quantized measurements up to and including t_{i-1}

K_i = Kalman Filter gain matrix at t_i

P_i = conditional covariance (under A1) of estimate had the ith measurement been linear

E_i = conditional covariance (under A1) of x_i given quantized measurements up to and including t_i

Under assumption A1 (28) and (29) correctly describe the propogation of the first two moments of the conditional distribution although it is no longer Gaussian. The Gaussian Fit Algorithm assumes that A1 is again true at time t_{i+1}, i.e., it "fits" a Gaussian distribution to the moments given by (28) and (29). To give some justification for this procedure, let $e=x-\hat{x}$ and subtract (28) from (10)

$$e_{i+1|i}=\Phi_i e_{i|i}+w_i \tag{30}$$

Since $e_{i|i}$ is not Gaussian, $e_{i+1|i}$ is not Gaussian either, although $e_{i+1|i}$ should tend toward a Gaussian distribution in the majority of cases because of the addition of Gaussian process noise w_i, and the mixing of the components of $e_{i|i}$ by the state transition matrix.

Because assumption A1 is not exact, the Gaussian Fit Algorithm described by the recursion relations (24-29) yields only approximations to the conditional moments. These recursion relations are very much like the Kalman Filter with two important differences:

1) The conditional mean of the measurement vector at t_i is used as the filter input. This conditional mean is computed on the assumption that the distribution of the measurement is $N[H_i\hat{x}_{i|i-1}, H_i M_i H_i^T+R_i]$

2) The conditional covariance equation (27) is being forced by the random variable $\mathrm{cov}(z_i|z_i \epsilon A_i)$. In general there is no steady state mean square error

for stationary input processes, and the filter weights are "random" until the previous measurement has been taken.

The primary advantages of the Gaussian Fit Algorithm are that 1) it is relatively easy to compute; 2) it can handle nonstationary data as easily as stationary data; and 3) its general operation is independent of the quantization scheme used. The primary disadvantages are that first, it requires more computation than the optimum linear filter and, second, it can be applied with some justification only to Gauss-Markov processes.

A recursive smoothing algorithm, which combines the output of two Gaussian Fit filters, is described in [17]. Like the Gaussian Fit Algorithm itself, the use of the smoothing technique is not limited to quantized measurements, but may be used with other nonlinearities.

Ensemble Average Performance Estimate for the Gaussian Fit Algorithm

The difficulty in analyzing the ensemble average performance of the Gaussian Fit Algorithm arises because the filter weights are random variables since they are functions of past measurements. Although one result of the filter computations is an approximation to the conditional covariance, this is not the ensemble covariance which is obtained by averaging over all possible measurement sequences. (For linear measurements, however, these two covariances are the same.) Approximate performance estimates are derived in the Appendix for the three systems considered in the next Section.

IV. PCM, PREDICTIVE QUANTIZATION, AND DATA COMPRESSION

PCM

The noiseless channel version of the PCM problem is shown in Fig. 1. Note that the quantizer output is the interval A_n in which the sample z_n falls.

When the quantum intervals are small enough, the conditional mean receiver for Gaussian variables is of the form (7) with the conditional mean of the measurement history (approximately) given by (22). In these equations z is that portion of $\{z_n\}$ upon which the estimate is based.

If $\{z_n\}$ is given by (10-15) the Gaussian Fit Algorithm may be applied in a straightforward manner to approximate the conditional mean receiver for the filtering and prediction problems. The smoothing problem is more complex and is considered in [17].

An estimate of the ensemble performance of the Gaussian Fit Algorithm in the PCM mode is given in the Appendix.

Predictive Quantization

System Description - Fig. 2 shows the noiseless channel version of the predictive quantization problem for a mean square error criterion. This system configuration is not as general as the one considered by Fine [9] since the quantizer is time invariant. The scalar random process $\{z_n\}$ is assumed to be the output of a system described by (10-15). The N-level quantizer is chosen beforehand, but is fixed once the system is in operation. The scalar feedback function $L_n(A_{n-1}, A_{n-2}, \ldots)$ is subtracted from the incoming sample z_n to minimize the mean square reconstruction error. The determination of L_n and the design of the quantizer are considered next.

System Optimization - Fine [9] outlines the system design procedure in three steps: 1) find the optimum receiver for a given transmitter; 2) find the optimum transmitter for a given receiver; 3) solve the simultaneous conditions of 1) and 2) for the optimum system. These are necessary, but not sufficient conditions [9]. Here we have already performed step 1) for the quadratic criterion since the conditional mean receiver is indicated in Fig. 2. Step 2) and step 3) are performed by choosing the optimum feedback quantity, L_n^o, such that

$$\mathcal{E}[\mathrm{cov}(x_n|L_n,A_n,A_{n-1},\ldots)|A_{n-1},\ldots]-\mathcal{E}[\mathrm{cov}(x_n|L_n^o,A_n,A_{n-1},\ldots)|A_{n-1},\ldots] \geq 0 \tag{31}$$

where cov $(x_n|\cdot)$ is the conditional covariance of the estimate, and the matrix inequality of (31) implies that the left-hand side is positive semidefinite. Choosing L_n in this manner assures a minimum variance estimate for any linear combination of the state variables.

An approximate solution to (31) can be found with the aid of the Gaussian Fit Algorithm. Under this assumption, the conditional moments are given in recursive form by (24-29), and the conditional covariance by (27) with the index i replaced by n. Let the quantized variable be u_n (Fig. 2), and furthermore let

$$\begin{aligned} u_n &= z_n-L_n(A_{n-1},\ldots) \\ u_n^o &= z_n-L_n^o(A_{n-1},\ldots) \end{aligned} \tag{32}$$

Note that in (27) we may use $\mathrm{cov}(u_n|\cdot)$ in place of $\mathrm{cov}(z_n|\cdot)$ since the two differ only by the constant L_n. If the N quantum intervals are denoted by $\{A^j,\ j=1,\ldots,N\}$, then substituting (27) into (31) and simplifying produces the scalar equation

$$\sum_{j=1}^{N} \mathrm{cov}(u_n|u_n \epsilon A^j)P(u_n \epsilon A^j)-\sum_{j=1}^{N} \mathrm{cov}(u_n^o|u_n^o \epsilon A^j)P(u_n^o \epsilon A^j) \geq 0 \tag{33}$$

where these quantities are computed by (17) and (18) under the assumption that z_n (hence u_n and u_n^o) are normally distributed.

By symmetry arguments it can be concluded that <u>for a well designed quantizer</u> L_n^o should be the (approximate) conditional mean of z_n,

$$L_n^o(A_{n-1},\ldots) = \mathcal{E}(z_n|A_{n-1},\ldots) = H_n\hat{x}_{n|n-1} \tag{34}$$

If, however, the conditional standard deviation of the prediction of z_n is very much smaller than the quantum interval widths, then the quantizer appears to be N-1 binary quantizers placed end to end. In this case L_n^o will probably be chosen to place the mean of u_n^o at one of these quantizer switch points. The quantizer design problem is considered in more detail in the Appendix.

<u>Computation of L_n</u> - It will be assumed that the quantum intervals are small enough so that the optimum choice of L_n is the conditional mean of z_n based on measurements at $\{t_{n-1},t_{n-2},\ldots\}$. The feedback function L_n may be based on a limited memory or a growing memory. The application of the Gaussian Fit Algorithm to the limited memory case is lengthy and will not be considered at this time.

The feedback function for the growing memory version uses the Gaussian Fit Algorithm on stationary or nonstationary processes from the first measurement to the most recent. Although the <u>memory</u> is increasing, the <u>storage</u> required to implement (24-29) remains constant. One of the great advantages of the Gaussian Fit Algorithm is that its operation is determined by the parameters of the input process, and it is not an iterively determined function as are most feedback operations of this type. The ensemble mean square error of this system is described in the Appendix.

There is one special case where the operation of the Gaussian Fit Algorithm becomes stationary and the feedback function and receiver become linear. When a binary quantizer is used, the filter weights as given by (25-27, 29) are predetermined functions of time because $\mathrm{cov}(z_i|\cdot)$ does not depend on the measurements. For a stationary input process the filter weights will approach a constant as more measurements are incorporated.

From (17),(24),(25),(28) and (34), the state space form of the stationary equations is

$$\hat{x}_{n|n-1} = \Phi\hat{x}_{n-1|n-1} \tag{35}$$

$$L_n = H\hat{x}_{n|n-1} \tag{36}$$

$$\hat{x}_{n|n} = \hat{x}_{n|n-1} + K_\infty \hat{r}_n \tag{37}$$

$$\hat{r}_n = \sqrt{\frac{2}{\pi}}\, \sigma_{z_\infty} \operatorname{sgn}(z_n - L_n) \tag{38}$$

$$\sigma_{z_\infty}^{\ 2} = HM_\infty H^T + R \tag{39}$$

$$K_\infty = M_\infty H^T (HM_\infty H^T + R)^{-1} \tag{40}$$

where M_∞ is the steady state solution to the covariance equations (26,27,29).

Data Compression

<u>System Description</u> - Fig. 3 shows the block diagram of the predictive-comparison type of data compression system. The analysis contained here is concerned only with the prediction and filtering aspects, and such important problems as buffer control, timing information, and channel noise are not considered.

The threshold device in Fig. 3 is a quantizer (one large quantum interval, many small quantum intervals), and the linear slope indicates that quantization during encoding can be neglected. The quantizer output is fed back through $L_n(A_{n-1},\ldots)$ and subtracted from the input z_n. If the magnitude of the difference, $|u_n|$, is less that α (a known parameter) then nothing is sent to the receiver; if $|u_n| > \alpha$, then u_n is sent to the receiver. If the receiver does a parallel computation of $L_n(A_{n-1},\ldots)$, then the system input z_n is calculated by adding u_n and L_n. The estimate of z_n when u_n is not sent may be performed under a variety of criteria; in all cases the error in the estimate is known to be less than α.

System Optimization - The criterion for the optimum system considered here will not contain data fidelity since this can be controlled through the choice of the threshold width. Instead, $L_n(A_{n-1},\ldots)$ is chosen solely on the basis that it shall minimize the (conditional) probability that u_n is sent to the receiver. This is not necessarily the same as minimizing the average number of samples sent out of the total number processed. This latter problem may be formulated as an optimal stochastic control problem requiring a dynamic programming approach. The solution would hardly be worth the effort.

The necessary equations for optimality are treated next. Let $y_{n-1}=\{z_{n-1},z_{n-2},z_{n-3},\ldots\}$ be the measurements used in determining L_n, and let $A'_{n-1}=\{A_{n-1},A_{n-2},\ldots\}$ be the region in which they fall. Then the probability of rejecting (not sending) the nth sample is

$$P(\text{reject}) = \int_{L_n-\alpha}^{L_n+\alpha} p_{z_n|y_{n-1}\epsilon A'_{n-1}}(\zeta_n)\,d\zeta_n \,, \tag{41}$$

where

$$p_{z_n|y_{n-1}\epsilon A'_{n-1}}(\zeta_n) = \frac{\displaystyle\int_{A'_{n-1}} p_{z_n,y_{n-1}}(\zeta_n,\eta_{n-1})\,d\eta_{n-1}}{\displaystyle\int_{A'_{n-1}} p_{y_{n-1}}(\eta_{n-1})\,d\eta_{n-1}} \tag{42}$$

The necessary condition that the probability of rejection be stationary with respect to L_n at $L_n^{\,o}$ is

$$p_{z_n|y_{n-1}\epsilon A'_{n-1}}(L_n^{\,o}+\alpha_n) = p_{z_n|y_{n-1}\epsilon A'_{n-1}}(L_n^{\,o}-\alpha_n) \tag{43}$$

or, substituting (42) into (43) provides

$$\int_{A_{n-1}^{\prime}} p_{z_n, y_{n-1}} (L_n^{o}+\alpha_n, \eta_{n-1}) d\eta_{n-1} = \int_{A_{n-1}^{\prime}} p_{z_n, y_{n-1}} (L_n^{o}-\alpha_n, \eta_{n-1}) d\eta_{n-1} \tag{44}$$

The above expression (44) does not require that $\{z_n\}$ be derived from a Markov process. If the threshold halfwidth α is small enough, (44) can be solved by power series approximations. Neglecting terms of third and higher order it may be verified that (44) reduces to

$$\frac{\partial}{\partial \zeta_n} \left[p_{z_n, y_{n-1}} (\zeta_n, \gamma_{n-1}) \right] \Bigg|_{L_n^{o}} = 0 \tag{45}$$

where γ_{n-1} is the collection of midpoints of the quantum intervals $\{A_{n-1}, A_{n-2}, \ldots\}$ Note that (45) is equivalent to

$$\frac{\partial}{\partial \zeta_n} \left[p_{z_n | y_{n-1}} (\zeta_n, \gamma_{n-1}) \right] \Bigg|_{L_n^{o}} = 0 \tag{46}$$

which means that, to the second order approximation, the optimum L_n is the mode of the density function of z_n conditioned on unquantized measurements γ_{n-1}. If $\{z_n\}$ are Gaussian random variables, then the conditional mode is the conditional mean and L_n^{o} is a linear operation.

When the $\{z_n\}$ process is generated by (10-15), the Gaussian Fit Algorithm may be used in the feedback path for arbitrarily wide thresholds. Regardless of the number of samples that have been rejected or sent, the distribution of z_n conditioned on quantized measurements at times $\{t_{n-1}, \ldots\}$ is assumed to be normal. The conditional probability of rejecting z_n is maximized by choosing L_n to be the (approximate) conditional mean, $H_n \hat{x}_{n|n-1}$.

The feedback function $L_n^o = \mathcal{E}(z_n | A_{n-1}, ..)$ is computed with the Gaussian Fit Algorithm. Observe that $\mathcal{E}(z_n | z_n \epsilon A_n)$ is just L_n if the sample falls within the threshold and is z_n if the sample is not quantized. The ensemble performance estimate of the system using the Gaussian Fit Algorithm is derived in the Appendix.

V. SIMULATION RESULTS

This section describes the results of digital computer simulations of the Gaussian Fit Algorithm as applied to the PCM, predictive quantization, and data compression systems described in Section IV. Bello, Lincoln, and Gish [10] present simulation results for predictive quantization with a binary quantizer. Their approach is a numerical approximation (by Monte Carlo techniques) to the optimum feedback function, whereas an analytical approximation (the Gaussian Fit Algorithm) is used here. They consider various memory lengths and a binary quantizer, and here we use a growing memory (finite storage) and arbitrary quantizers. Although the Gaussian Fit Algorithm and its performance estimate may be used on nonstationary data, only stationary data have been simulated as yet.

Simulation Description

Input Process- The simulated second order Gauss-Markov input process is the sampled output of a linear system driven by Gaussian white noise. The transfer function of the shaping filter is the same as used in [10]

$$H(s) = \frac{c}{(1+\tau s)^2} \tag{47}$$

where the gain c is chosen to provide the proper variance at the output. Observation noise was not used here, but is considered in [17]. Thus the autocorrelation of the input process is

$$\phi_{zz}(n) = \mathcal{E}(z_i z_{i+n}) = (1+|n|/r)\exp(-|n|/r) \tag{48}$$

where

$r = \tau/T$ = number of samples per time constant τ

T = time between samples

Error Measurement - Each system was simulated by operating on 5000 consecutive samples. The estimation errors were squared and averaged to give an estimate of the ensemble mean square error of the system. The autocorrelations of the estimation errors were measured and from this the confidence limits have been assessed as being greater than a 90 percent probability that the measured error variance lies within 10 percent of its true value.

PCM and Predictive Quantization

Fig. 4 displays the ratio of signal variance to ensemble mean square estimation error (expressed in decibels) as a function of the number of quantizer quantum intervals. Both the PCM and predictive quantization systems are shown with the input process parameter $r=2.5$. The lines are the performance estimates (as derived in the Appendix) and the data points are the simulation results. The predictive quantization system performs significantly better than the PCM system, as is to be expected. The performance estimate is quite accurate except for PCM with a small number of quantum intervals (less than 5). Here the estimate is optimistic, a characteristic that has been noted in other simulations [17]. The quantizer quantum intervals have been chosen to minimize the ensemble mean square error (see Appendix).

Fig. 5 shows how the predictive quantization system with a binary quantizer reacts to different signal correlations. The performance estimate and the simulation results are exhibited as a function of the input process parameter r. (As a point of reference, the adjacent sample correlation is 0.736 for $r=1$; 0.938 for $r=2.5$; 0.9988 for $r=20$.) Again, the performance estimate is quite accurate.

Data Compression

Fig. 6 contains the outcomes for the predictive-comparison data compression system. Performance estimates and simulation results of the mean square error and sample compression ratio are shown as a function of (α/σ_z^a), the ratio of threshold halfwidth to a priori standard deviation. Note the excellent agreement between performance estimates and simulation results.

VI. CONCLUSIONS

Expectations conditioned on quantized measurements ($z \epsilon A$) can be found in two steps: 1) Find the expectation conditioned on z; 2) Average this (conditional) expectation conditioned on $z \epsilon A$. When applied to discrete time linear-Gaussian systems, it was shown that the conditional mean of the system's state vector can be found without Bayes' Rule by passing the conditional mean of the measurement history through the Kalman filter. This result provides one common point of departure for system design.

Two nonlinear approximations are considered for Gaussian variables. The first uses a power series expansion and neglects fourth and higher order powers of the quantum-interval-to-standard-deviation ratio. The second approach, called the Gaussian Fit Algorithm, assumes that the conditional distribution is normal. It is a recursive computation for arbitrarily wide quantum intervals, but is limited to nth order Gauss-Markov processes.

The approximations are applied to the noiseless channel versions of three digital systems: PCM, predictive quantization, and predictive-comparison data compression. Both methods can be used on stationary and nonstationary data, and can be used in the feedback path without additional calculations e.g., Monte Carlo. The Gaussian Fit Algorithm uses a growing memory (but finite storage) for these computations. Estimates of the ensemble mean square reconstruction error are derived for the Gaussian Fit Algorithm when used in each of the three systems. Simulation results indicate that these ensemble performance estimates are quite accurate (except for very coarse PCM), so that parametric studies with Monte Carlo techniques are not required to evaluate the system's ensemble mean square error.

PRECEDING PAGE BLANK NOT FILMED.

APPENDIX

PERFORMANCE ESTIMATES FOR THE GAUSSIAN FIT ALGORITHM

This Appendix considers the performance of the Gaussian Fit Algorithm over the ensemble of time functions for which it is intended. An approximation to the ensemble mean square error is found for stationary and nonstationary data, and the analysis is applied to the three digital systems of Section IV.

General Form of the Performance Estimate

The Gaussian hypothesis for the Gaussian Fit Algorithm is assumed so that (24-29) describe the propagation of the first two moments. The ensemble average covariance is found by averaging the conditional covariance over all possible measurement sequences. Consider M_{n+1}, the conditional covariance just prior to the measurement at t_{n+1}. From (25-27) it is seen that (29) can be written

$$M_{n+1} = \Phi_n P_n(M_n)\Phi_n^T + Q_n + \Phi_n K_n(M_n)\operatorname{cov}(z_n|A_n, M_n, \hat{x}_{n|n-1})K_n^T(M_n)\Phi_n^T \tag{A.1}$$

where the conditional mean and covariance of z_n just prior to the reading at t_n are explicitly shown in $\operatorname{cov}(z_n|\cdot)$. Let M^*_{n+1} be the ensemble average of M_{n+1} and formally take the ensemble expectation of both sides of (A.1)

$$M^*_{n+1} = \int dM_n\, p(M_n)[\Phi_n P_n(M_n)\Phi_n^T + Q_n + \Phi_n K_n(M_n)W(M_n)K_n^T(M_n)\Phi_n^T] \tag{A.2}$$

Here $p(M_n)$ is the ensemble probability density function for the matrix M_n, and the matrix $W(M_n)$ is

$$W(M_n) = \mathcal{E}[\operatorname{cov}(z_n|A_n, \hat{x}_{n|n-1}, M_n)|M_n] \tag{A.3}$$

The form of $W(M_n)$ depends on the quantization scheme and explicit forms are given below for three types of digital systems. The case of a scalar z_n is of special interest. Let $\{A^j, j=1,\ldots,N\}$ denote the N quantum intervals, and let

$$\bar{z}_n = H_n \hat{x}_{n|n-1} \tag{A.4}$$

$$\sigma_{z_n}^2 = H_n M_n H_n^T + R_n$$

Then the expression for the scalar $W(M_n)$ becomes

$$W(M_n) = \mathcal{E}\left[\sum_{j=1}^{N} \operatorname{cov}(z_n | z_n \epsilon A^j, \bar{z}_n, \sigma_{z_n}^2) \cdot P(z_n \epsilon A^j | \bar{z}_n, \sigma_{z_n}^2) \middle| \sigma_{z_n}^2\right] \tag{A.5}$$

Regardless of the form of $W(M_n)$, an approximate solution to (A.2) can be found by expanding the right-hand side in a Taylor series about the ensemble mean, M_n^*, and neglecting second and higher order terms. This will be an accurate solution when there is small probability that M_n is "far" from M_n^*, and results in the equation

$$M_{n+1}^* \approx \Phi_n P_n(M_n^*)\Phi_n^T + Q_n + \Phi_n K_n(M_n^*) W(M_n^*) K_n^T(M_n^*)\Phi_n^T \tag{A.6}$$

with

$$M_1^* = \Phi_0 P_0 \Phi_0^T + Q_0$$

By similar reasoning the ensemble average covariance just after the $(n+1)^{st}$ measurement, E_{n+1}^*, is given by

$$E_{n+1}^* \approx P_{n+1}(M_{n+1}^*) + K_{n+1}(M_{n+1}^*) W(M_{n+1}^*) K_{n+1}^T(M_{n+1}^*) \tag{A.7}$$

PCM

Computation of W(M*) - The estimate of the ensemble mean square error of the Gaussian Fit Algorithm in the PCM mode is determined by (A.6) with W(M) determined as follows. Dropping the subscripts, let the N quantum intervals $\{A^j\}$ be defined by the boundaries $\{d^j, j=1,...N+1\}$, that is

$$z \epsilon A^j \text{ if } d^j \leq z < d^{j+1} \quad j=1,2,...,N \tag{A.8}$$

A quantizer distortion function is defined for a Gaussian variable and a particular quantizer

$$D\left(\frac{\bar{z}}{\sigma_z}, \frac{d}{\sigma_z}\right) = \frac{1}{\sigma_z^2} \sum_{j=1}^{N} \text{cov}(z|z\epsilon A^j, \bar{z}, \sigma_z^2) P(z\epsilon A^j|\bar{z}, \sigma_z^2) \tag{A.9}$$

where the conditional covariance and probability are evaluated by (17) and (18). This distortion function is the normalized minimum mean square error in reconstructing the quantizer input. Substituting (A.9) into (A.5) provides

$$W(M) = \sigma_z^2 \int D(\frac{\bar{z}}{\sigma_z}, \frac{d}{\sigma_z}) p(\bar{z}|M) d\bar{z} \tag{A.10}$$

An approximation to the ensemble statistics of $\bar{z}$ conditioned on M will be considered next. Let $\bar{x}$ be the mean of the state conditioned on past data Z. Then from (A.4)

$$\begin{aligned} \bar{z}^2 &= H\bar{x}\bar{x}^T H^T \\ &= H[\mathcal{E}(xx^T|Z) - \text{cov}(x|Z)]H^T \end{aligned} \tag{A.11}$$

The ensemble average of $\bar{z}$ is zero (or can be made zero by change of variable) so that averaging (A.11) over all possible measurements Z yields the variance of $\bar{z}$, $\sigma_{\bar{z}}^2$

$$\begin{aligned} \sigma_{\bar{z}}^2 &= H[M^a - M^*]H^T \\ &= (\sigma_z^a)^2 - (\sigma_z^*)^2 \end{aligned} \tag{A.12}$$

where M^a is the a priori covariance of the process and

$$(\sigma_z{}^a)^2 = HM^aH^T+R$$
$$(\sigma_z{}^*)^2 = HM^*H^T+R \qquad \text{(A.13)}$$

Now it is assumed that the distribution of $\bar{z}$ is close to $N[0,\sigma_{\bar{z}}{}^2]$ (this is strictly true when the measurements are linear), and the desired result is

$$W(M_n{}^*) = (\sigma_{z_n}{}^*)^2 \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}\left(\frac{\bar{z}}{\sigma_{\bar{z}_n}}\right)^2}}{\sqrt{2\pi}\,\sigma_{\bar{z}_n}} D\left(\frac{\bar{z}}{\sigma_{z_n}{}^*}, \frac{d}{\sigma_{z_n}{}^*}\right) d\bar{z} \qquad \text{(A.14)}$$

where the subscripted standard deviations are found from subscripted versions of (A.12) or (A.13). Although further approximations can be made to (A.14) it can be evaluated quite easily by numerical quadrature.

The recursive relations (A.6) and (A.7) may be used with (A.14) to obtain an approximate estimate of the ensemble mean square error for nonstationary data. For stationary data, $E_M{}^*$ becomes an approximation to the ensemble covariance of estimation errors using an estimate based on M measurements. When M is very large and $E^*_{M+1} \simeq E_M{}^*$, this becomes the performance estimate for the infinite memory estimator.

The Optimum Quantizer - For three or more quantum levels, the quantizer parameters $\{d^i\}$ may be chosen to minimize the ensemble covariance $E_M{}^*$. This is considered in [17], where it was found for stationary processes that $E_\infty{}^*$ is relatively insensitive to changes in $\{d^i\}$ if the quantizer is optimized for the a priori variance $(\sigma_z{}^a)^2$. The insensitivity is caused by two competing factors: 1) the filtering action reduces the variance implying that the quantum intervals should be reduced; 2) but from (A.12) a smaller $\sigma_z{}^*$ means a larger variance of the conditional mean and the quantum intervals should be increased to insure efficient quantization when $\bar{z}$ is away from zero.

Predictive Quantization

As with the PCM system, an estimate of the ensemble average mean square error is generated by the recursive solution of (A.6) and (A.7) with a different form for $W(M_n^*)$. Recall that the feedback function L_n is the mean of z_n conditioned on all past measurements. This has the effect of setting $\bar{z}=0$ in the distortion function given by (A.9). (Actually, u_n is being quantized, but u_n and z_n differ only by a constant so their covariances are the same.) Consequently, $W(M_n^*)$ for the predictive quantization becomes

$$W(M_n^*) = (\sigma_{z_n}^*)^2 D(0, \frac{d}{\sigma_{z_n}^*}) \tag{A.15}$$

Now (A.6),(A.7) and (A.15) describe the estimates of the ensemble covariance for the growing memory predictive quantization scheme.

The quantizer design for a stationary input may be performed as follows. Temporarily assume that the quantizer is time varying, and choose the parameters $\{d^i\}$ at time t_n to be optimum for $(\sigma_{z_n}^*)^2$ Now $D(0, \frac{d}{\sigma_{z_n}^*})$ becomes $D_g(N)$, the minimum distortion for a unit variance, zero mean Gaussian variable, and it is a function only of the number of quantum levels N. (See Max [22] for tabulations of $\{d^i\}$ and $D_g(N)$.) With this time varying quantizer, $W(M_n^*)$ becomes

$$W(M_n^*) = (\sigma_{z_n}^*)^2 D_g(N) \tag{A.16}$$

As n approaches infinity the ensemble covariance and thus the quantizer parameters approach a constant. This final quantizer minimizes the ensemble covariance because using any distortion other than $D_g(N)$ yields a larger solution to the Riccati equation.

Data Compression

The W(M) function for the data compression system is identical to predictive quantization except that there is only one quantum interval (of halfwidth α) in the distortion function. Substituting (17) and (18) into (A.15)

$$W(M_n^*) = (\sigma_{z_n}^*)^2 \left[\int_{-\frac{\alpha}{\sigma_{z_n}^*}}^{\frac{\alpha}{\sigma_{z_n}^*}} \frac{e^{-\frac{1}{2}u^2} du}{\sqrt{2\pi}} - 2\left(\frac{\alpha}{\sigma_{z_n}^*}\right) \frac{e^{-\frac{1}{2}\left(\frac{\alpha}{\sigma_{z_n}^*}\right)^2}}{\sqrt{2\pi}} \right] \tag{A.17}$$

Equations (A.6),(A.7) and (A.17) describe an approximation to the ensemble covariance, and M_∞^* is the steady state ensemble covariance of the prediction error. The sample compression ratio, CR, is a common figure of merit for data compression systems; it is the ratio of the number of input samples to the number of transmitted samples. For the system considered here it is

$$CR(M_\infty^*) \approx \left[1 - \int_{-\alpha/\sigma_{z_\infty}^*}^{\alpha/\sigma_{z_\infty}^*} \frac{e^{-\frac{1}{2}u^2} du}{\sqrt{2\pi}} \right]^{-1} \tag{A.18}$$

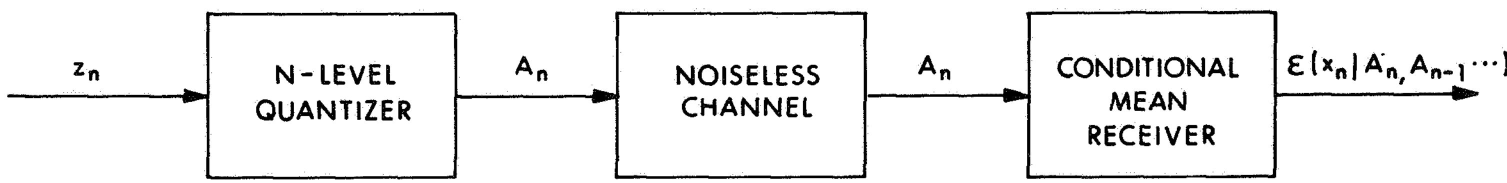

Fig. 1 Noiseless channel PCM.

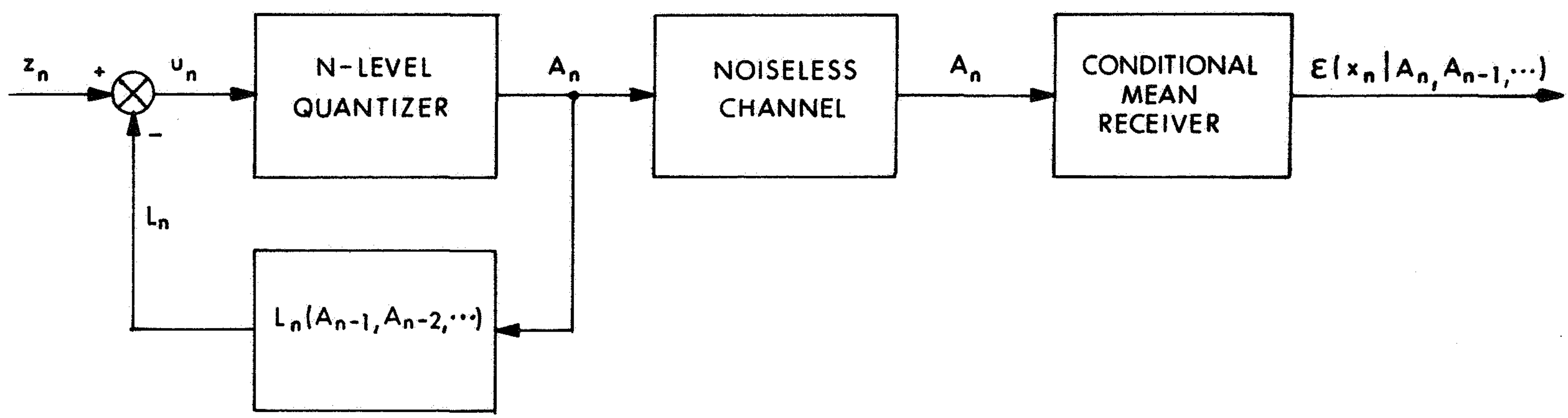

Fig. 2 Noiseless channel predictive quantization.

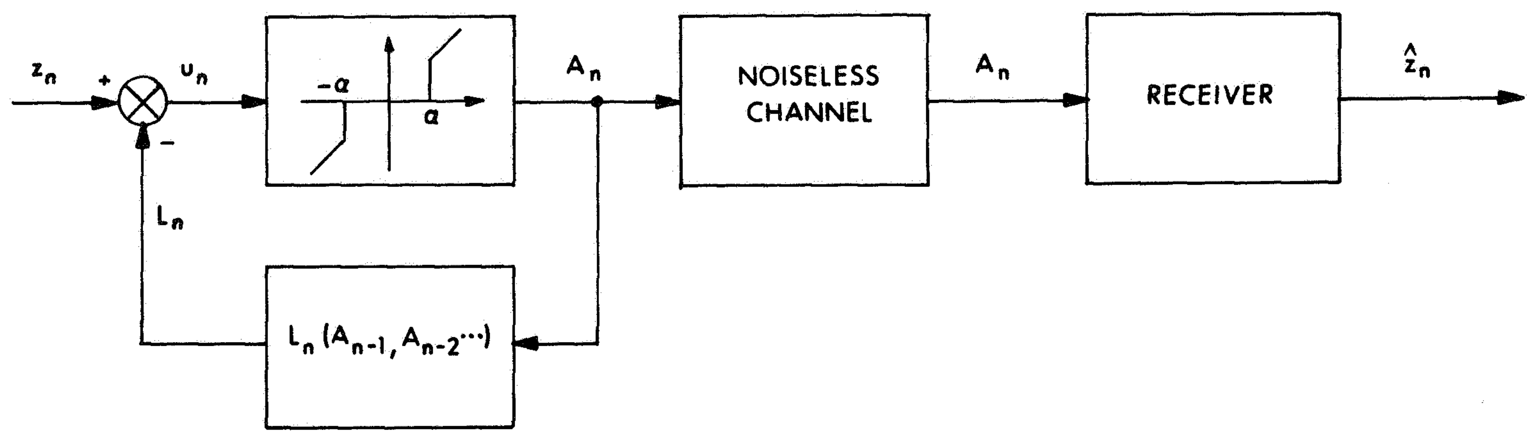

Fig. 3 Noiseless channel predictive-comparison data compression.

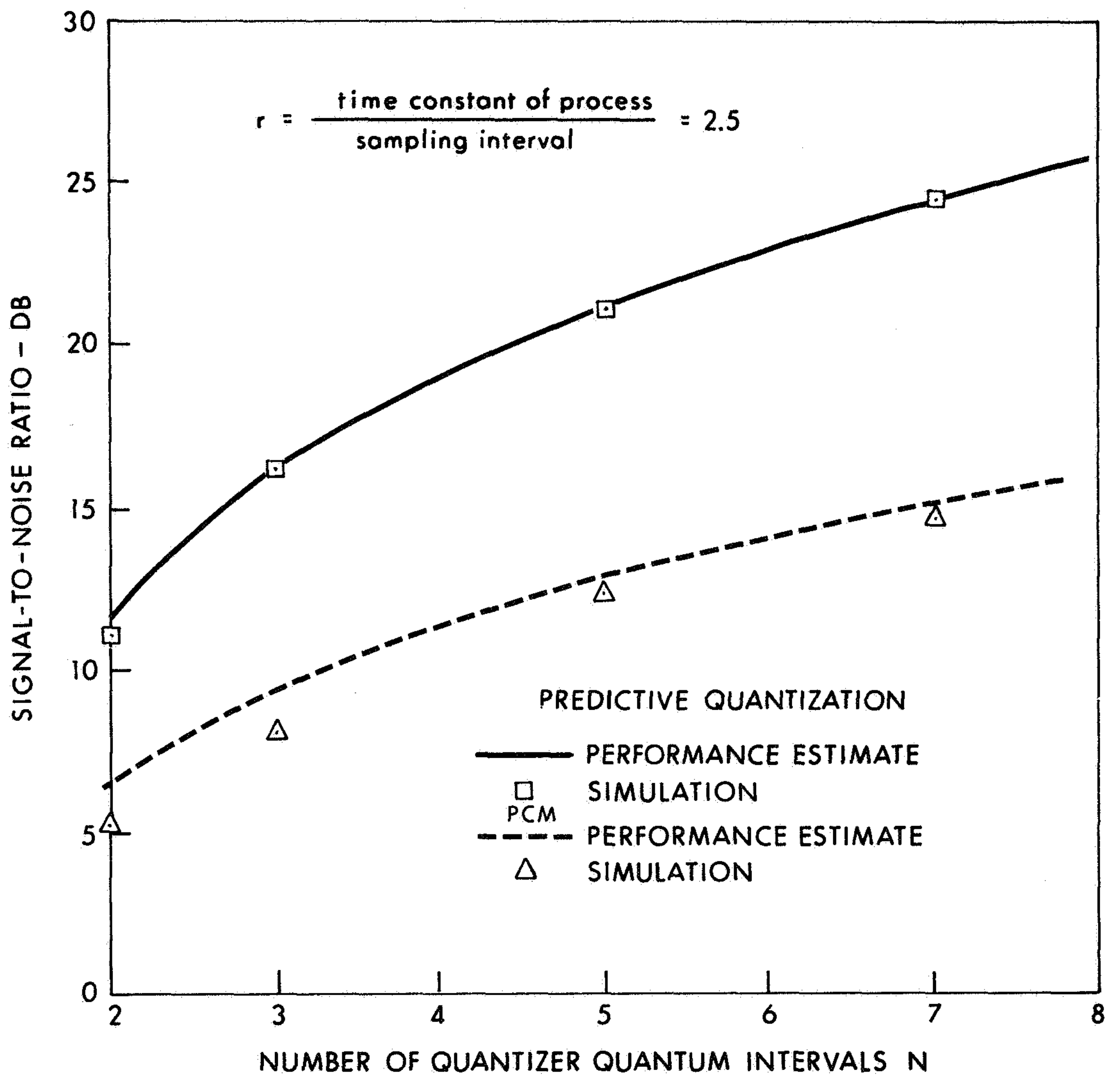

Fig. 4 SNR for the Gaussian Fit Algorithm, PCM and predictive quantization (r = 2.5).

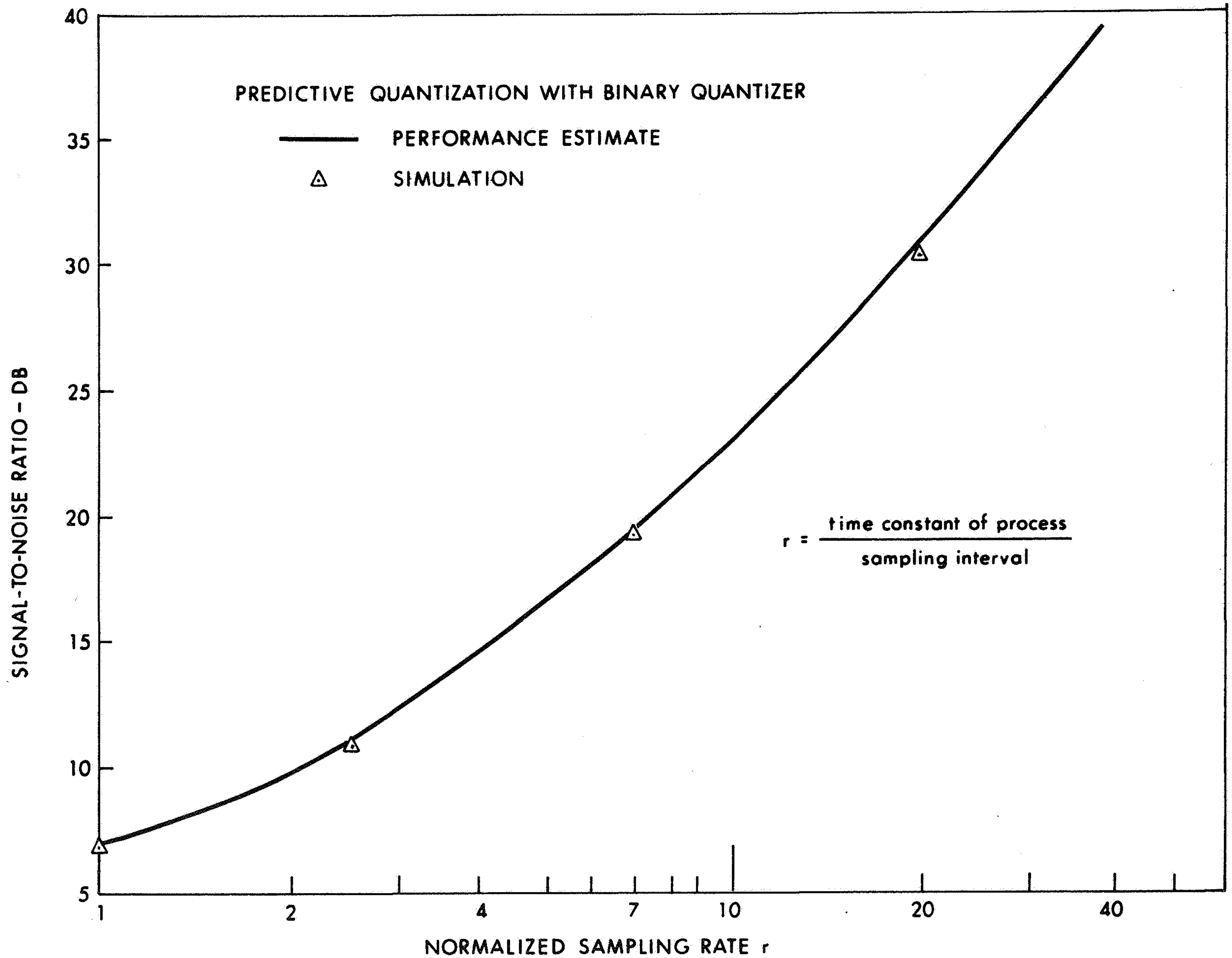

Fig. 5 SNR for the Gaussian Fit Algorithm, predictive quantization, binary quantizer.

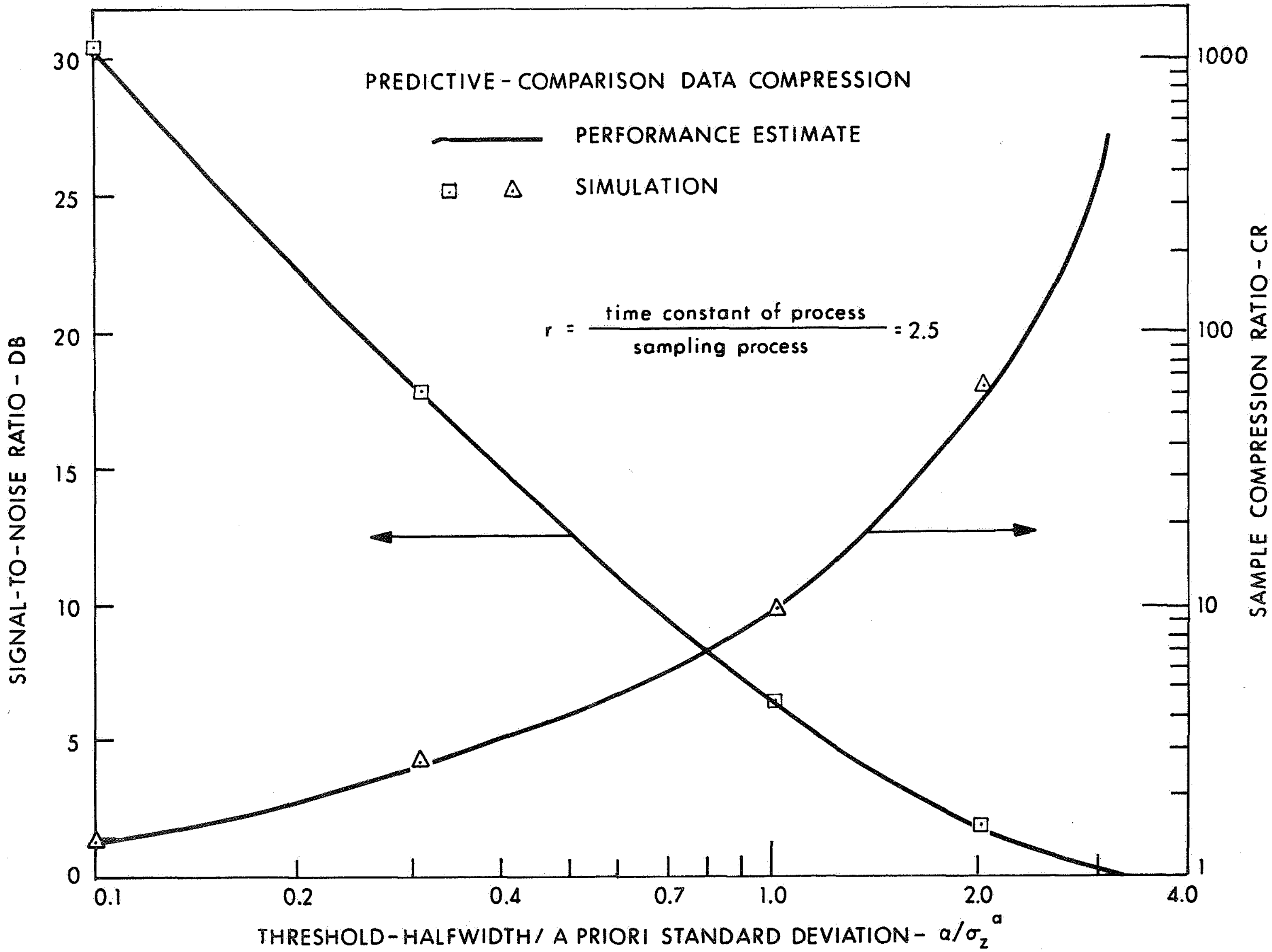

Fig. 6 SNR and sample compression ratio for the Gaussian Fit Algorithm, predictive-comparison data compression (r = 2.5).

REFERENCES

1. Proceedings of the IEEE, Special Issue on Redundancy Reduction, 55, March 1967.

2. Ruchkin, D., "Linear Reconstruction of Quantized and Sampled Signals", IRE Trans. on Communications Systems, Dec. 1961, pp. 350-355.

3. Steiglitz, K., "Transmission of an Analog Signal Over a Fixed Bit-Rate Channel", IEEE Trans. on Information Theory, IT-12, 1966, pp. 469-474.

4. Kellog, W., "Information Rates in Sampling and Quantization,", IEEE Trans. on Information Theory, IT-13, 1967, pp. 506-511.

5. Ho, Y.C., Lee, R.C.K., "A Bayseian Approach to Problems in Stochastic Estimation and Control", Proc. JACC, 1964, pp. 382-387.

6. Bucy, R., "Nonlinear Filtering Theory", IEEE Trans. on Automatic Control, AC-10, 1965, p. 198.

7. Wonham, W.M., "Some Applications of Stochastic Differential Equations to Nonlinear Filtering", J.SIAM Control, Ser.A, 2, 1964, pp. 347-369.

8. Meier, L., Korsak, A., Larson, R., "Effect of Data Quantization on Tracker Performance", Tech, Mem. 3, Project 6642, Stanford Research Institute, Menlo Park, Calif. Oct. 1967.

9. Fine, T., "Properties of an Optimum Digital System and Applications", IEEE Trans. on Information Theory, IT-10, 1964, pp. 287-296.

10. Bello, P., Lincoln, R., and Gish, H., "Statistical Delta Modulation", Proc. of IEEE, 55, March 1967, pp. 308-319.

11. Gish, H., "Optimum Quantization of Random Sequences", Rept. 529, Div. of Eng. and Applied Physics, Harvard Univ., May 1967.

12. O'Neal, J., "Predictive Quantizing Systems (Differential Pulse Code Modulation) for the Transmission of Television Signals", B.S.T.J., 45, 1966, pp. 689-720.

13. Irwin, J., and O'Neal, J., "The Design of Optimum DPCM (Differential Pulse Code Modulation) Encoding Systems Via the Kalman Predictor", Preprints JACC, June 1968, pp. 130-136.

14. Davisson, L., "An Approximate Theory of Prediction for Data Compression", IEEE Trans. on Information Theory, IT-13, April 1967, pp. 274-278.

15. Davisson, L., "Theory of Adaptive Data Compression", in Advances in Communication Systems, Balakrishnan, ed., Academic Press, New York, 1966.

16. Doob, J., Stochastic Processes, Wiley, New York, 1953.

17. Curry, R., "Estimation and Control with Quantized Measurements", Ph.D. Dissertation, Dept. of Aeronautics and Astronautics, MIT, May 1968. Also Report TE-23, Experimental Astronomy Laboratory, MIT, Cambridge, Mass.

18. Fraser, D., "A New Technique for the Optimal Smoothing of Data", Sc.D. Thesis, Department of Aeronautics and Astronautics, MIT, January, 1967.

19. Jazwinski, A., "Filtering for Nonlinear Dynamical Systems", IEEE Transactions on Automatic Control, AC-11, 1966, p. 765.

20. Bass, R., and Schwartz, L., "Extensions to Multichannel Nonlinear Filtering", Hughes Report SSD 60220R, Feb., 1966.

21. Fisher, J., "Conditional Probability Density Functions and Optimal Nonlinear Estimation", Ph.D. Dissertation, UCLA, Los Angeles, Calif., 1966.

22. Max, J., "Quantizing for Minimum Distortion", IRE Trans. on Information Theory, March 1960, pp. 7-12.